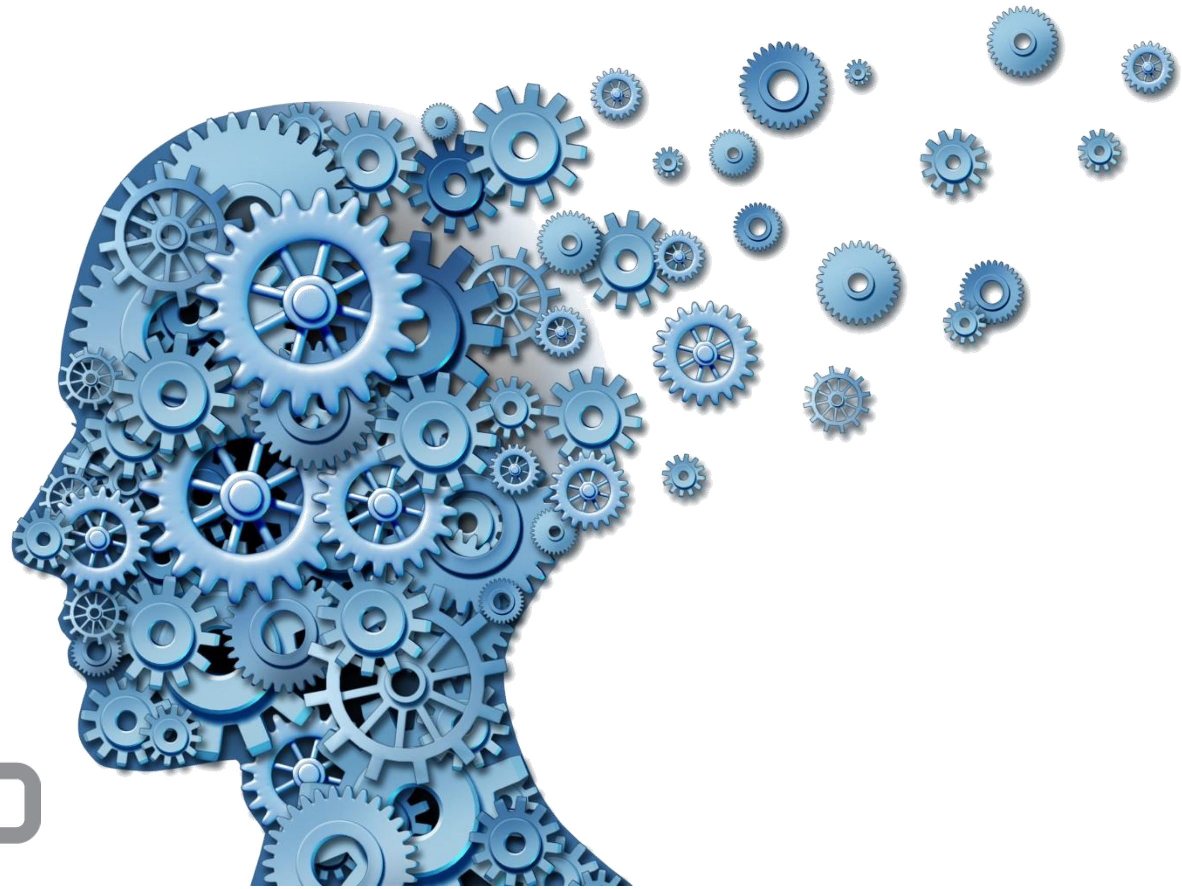
Zürcher Hochschule
für Angewandte Wissenschaften

Beyond ImageNet –
Deep Learning in Industrial Practice

4th Swiss Conference on Data Science, Bern, June 16, 2017

Thilo Stadelmann & Oliver Dürr



datalab
www.zhaw.ch/datalab

1

Deep Learning in a nutshell

Deep learning (DL) @ ZHAW Datalab

Initial group of 10+ researchers to start research line in 2014

- > 11 projects since, 9 of which with industry participation
  (19 month duration on average, >7M CHF overall volume, several publications)
- > 20 students in thesis projects per semester
  (Bachelor & Master level)
- 125k CHF investment in GPU resources up to fall 2017
- New modules in Bachelor, Master, and professional education curricula

Statistician*

Visionary*

* a very lossy compression of the truth…



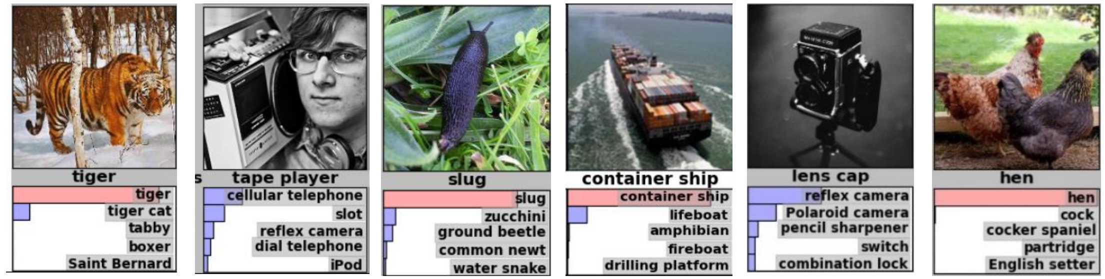


Marek Fernando Mark Jan Oliver Gabriel Olga Stefan Hans-Peter Melanie Benjamin Elvis Beate Thilo Vasily Lukas

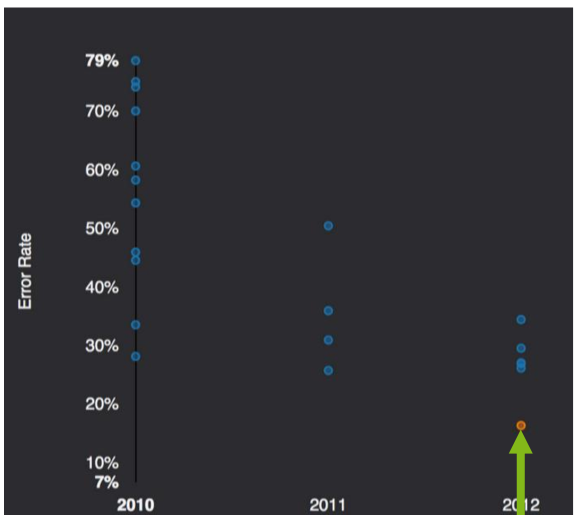
Zürcher Hochschule
für Angewandte Wissenschaften

DL History: ImageNet competition starts hype

1000 classes
1 Mio. samples



…



A. Krizhevsky uses a convolutional deep neural network (CNN) for the first time.

2015: computers *learned to see*

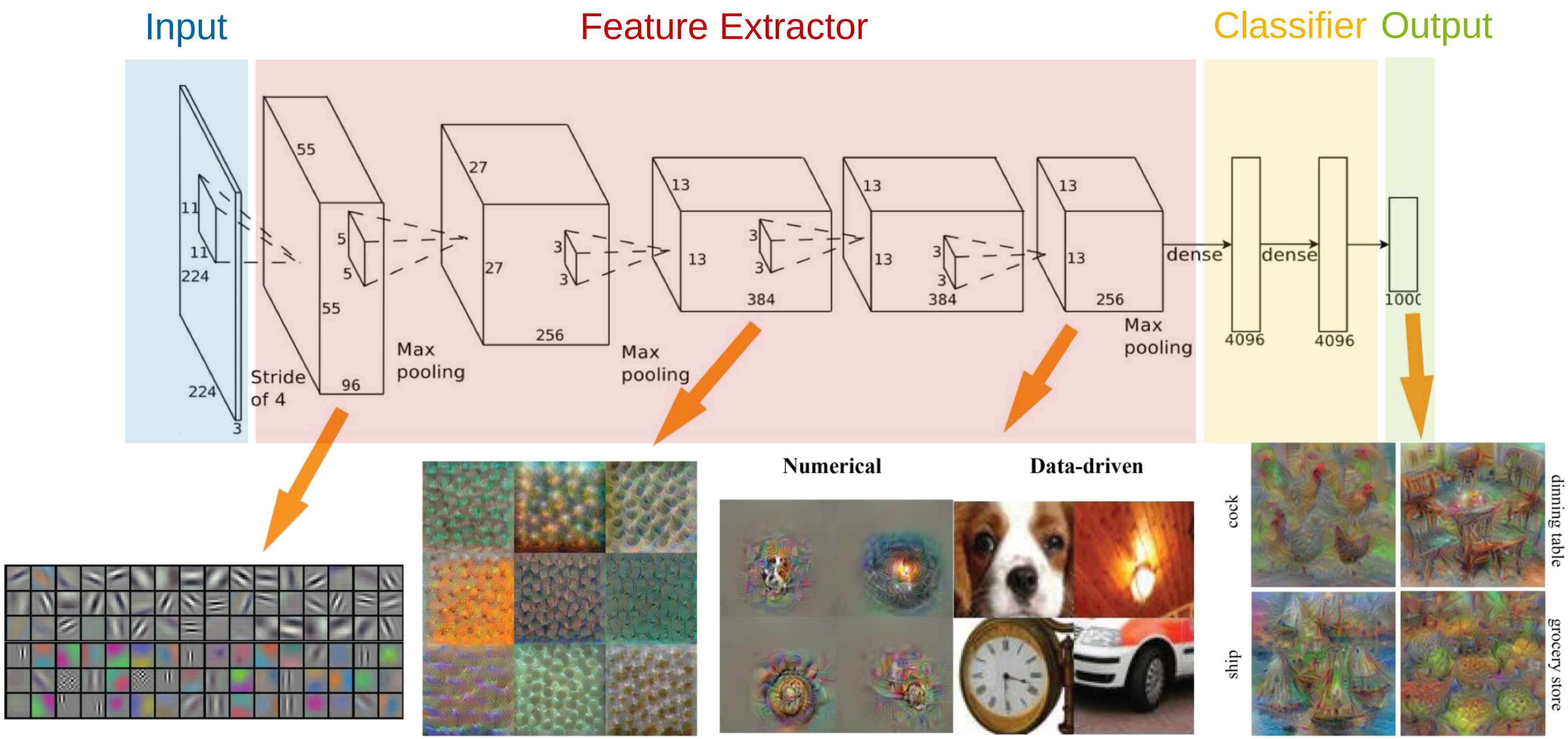
4.95% Microsoft (February 6, 2015)
→ surpassing human performance of 5.10%

4.80% Google (February 11, 2015)

4.58% Baidu (May 1, 2015)

3.57% Microsoft (December 10, 2015)

Convolutional Neural Networks
Building blocks forming a sophisticated architecture



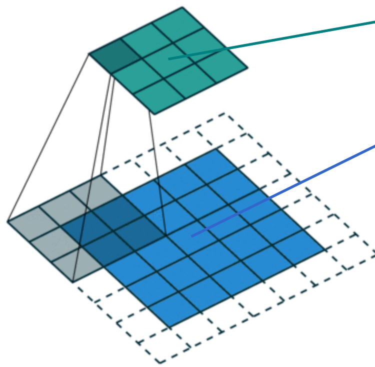
Conv 1: Edge+Blob | Conv 3: Texture | Conv 5: Object Parts | Fc8: Object Classes

Source: http://vision03.csail.mit.edu/cnn_art/data/single_layer.png

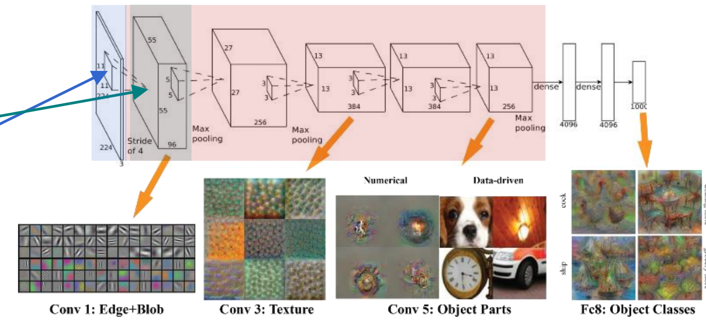
Micro building blocks: layers

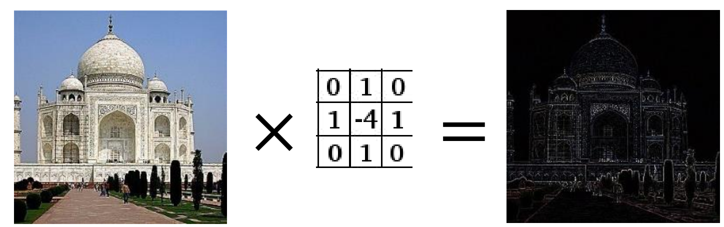
E.g., convolutional layers

- **Strong in finding local patterns** (i.e., 2D structure)
  → based on the idea of filtering:

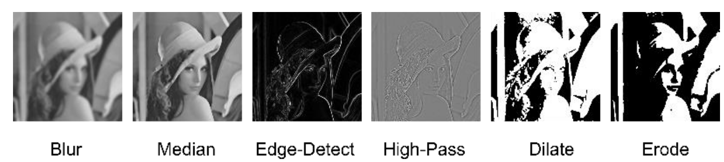


Dumoulin, Visin, «A guide to convolution arithmetic for deep learning », 2016





Based on https://docs.gimp.org/en/plug-in-convmatrix.html



http://blog.teledynedalsa.com/wp-stuff/uploads/2012/05/AGhannoum-FilterExamples.png

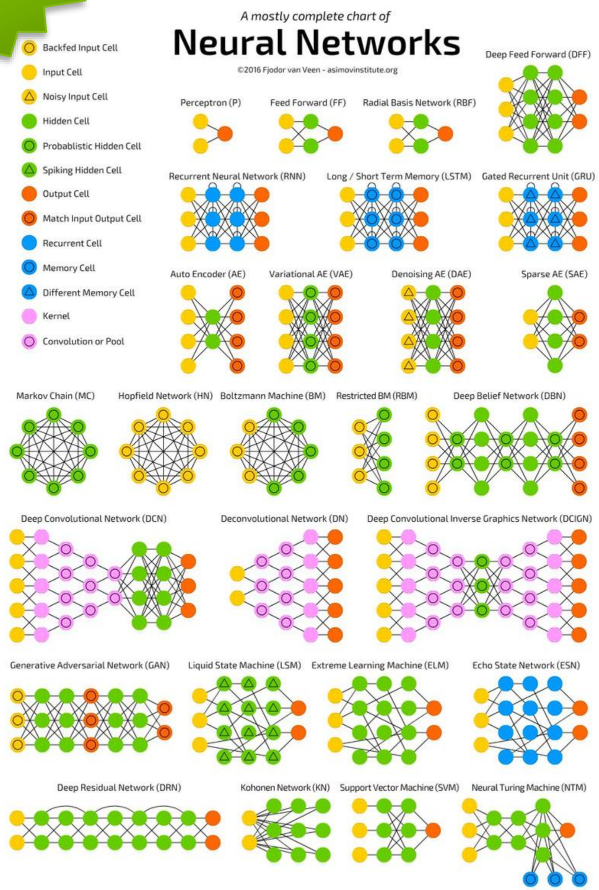
- Translation- and scale invariant
  → through down sampling (strided conv's or max pooling)
- Fast to compute
  → all possible filter locations share the 3×3 weights

Alternatives

- Recurrent layers, different output layers, …
  → for **temporal relationships**, regression, **distribution matching**, …

Macro building block: whole architectures

Example

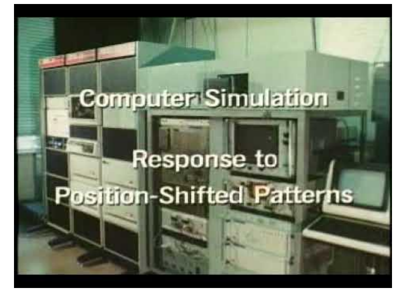


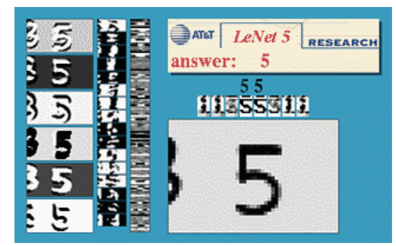
Source: http://www.asimovinstitute.org/neural-network-zoo/

What makes CNNs work?

More history

- Biological plausibility known since 1959 (Hubel & Wiesel)
- First CNN ("Neocognitron") in 1980 (Fukushima)
- Automatic bank check reading in 1998 (LeCun et al.)
- But: General breakthrough in computer vision only in 2012 (see above)



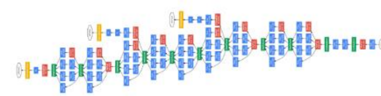


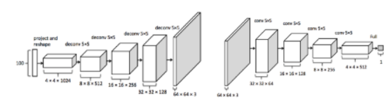
"…**most of this progress** is not just the result of more powerful hardware, larger datasets and bigger models, but **mainly a consequence of new ideas, algorithms** and **improved** network **architectures**".

Szegedy et al., "Going Deeper with Convolutions", 2014

What's different now?

- **Big Data** (labeled images via the web)
- **Compute power** (consumer GPUs, i.e. NVIDIA's GeForce series)
- **Algorithmic improvements**
  → Regularization: **Dropout**
  → Optimization: **ADADELTA**
  → Trainability ("gradient flow"): **Batchnorm**, **ReLU**
  → Exploitation of available data: **Augmentation**, **transfer learning**
  → More powerful architectures: ResNet, **Fully Convolutional NN**, Generative Adversarial N, YOLO, …





Introduction → Use Cases → Lessons Learned

2

From image classification to paths less travelled

Case Study I (Limited Resources)
Face Recognition on Raspberry Pi

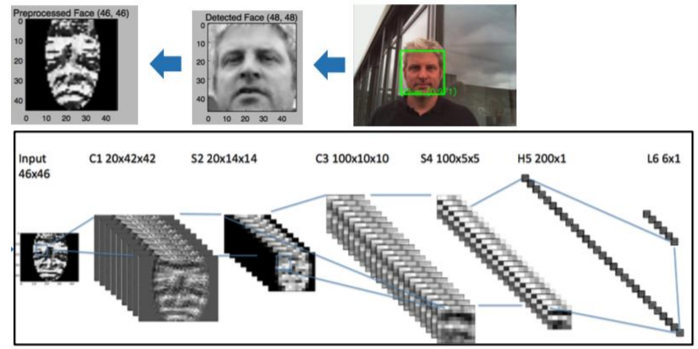
Architecture and Training Set

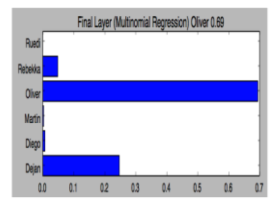
Training indoors

Approx. 40 images of 6 persons

Prediction done on Raspberry Pi

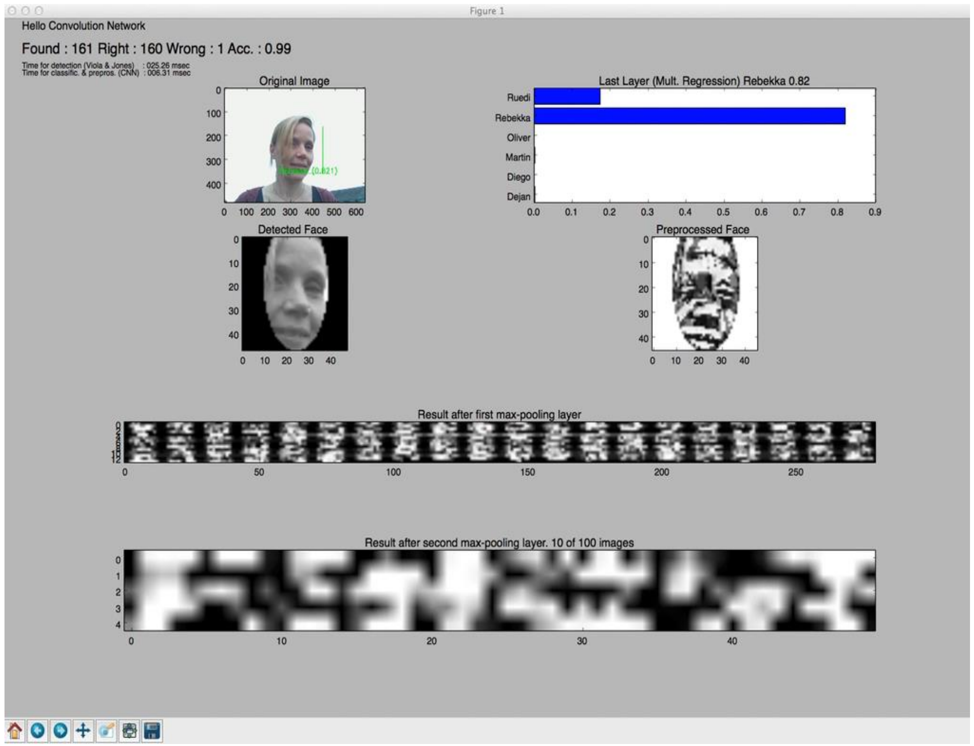
Testing outdoors



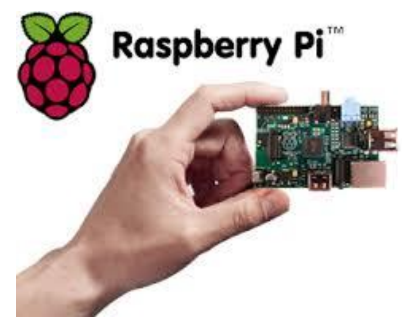


Forward pass

train on GPU…



…predict on Raspberry Pi



Results

Method	Accuracy	Classification Time [msec]	Enrollment Rate N_e/N	Total Time Per Face [msec]
CNN (p_0=0.85)	99.59%	105 +/- 8	250/278	529 +/- 64
CNN (p_0=0.00)	97.48%	105 +/- 8	278/278	529 +/- 64
Fisherfaces (no al.)	88.5%	54 +/- 11	278/278	511 +/- 89
Fisherfaces (al.)	96.87%	535 +/- 89	192/278	1006 +/- 18

- Faster then traditional pipeline
- No alignment needed

- Alternative implementation on Android (BA Thesis)
- Similar: BA Thesis for classification of weasel

Dürr, O., Pauchard, Y., Browarnik, D; Axthelm, R.; Loeser, M. (2015): *Deep Learning on a Raspberry Pi for Real Time Face Recognition*. EG 2015 – Posters 11-12.

Apodemus speciosus

Case Study II
HCS Screening

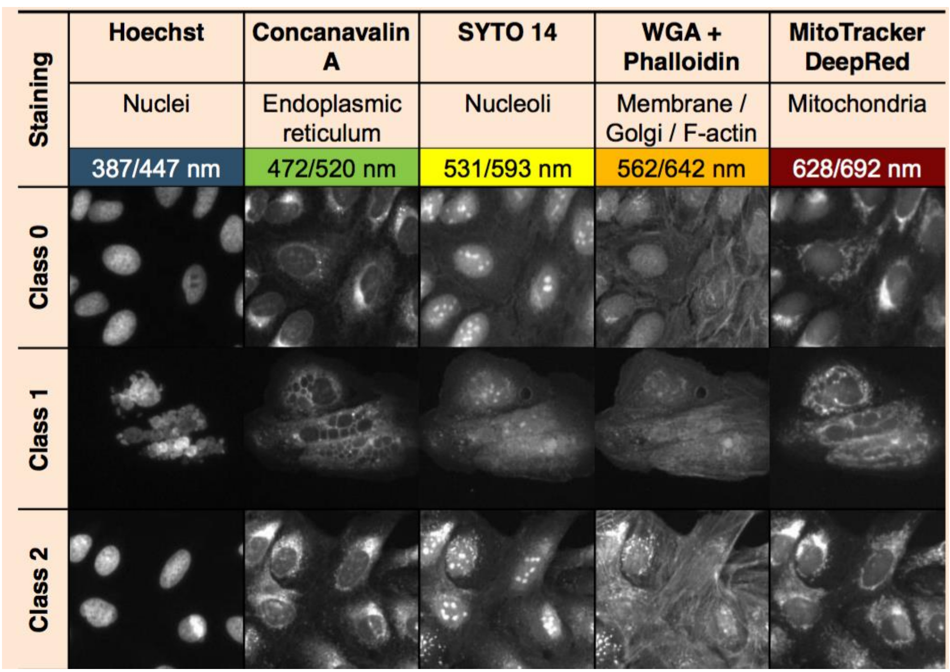
Phenotypic High Content Screening

- Compounds (potential drugs) are placed in a well (test-tube)
- Some compounds make cell to react in a particular way (change the phenotype)
- Interest in compounds which change the phenotype in a particular way

Phenotype 0

Phenotype 1

Phenotype 2

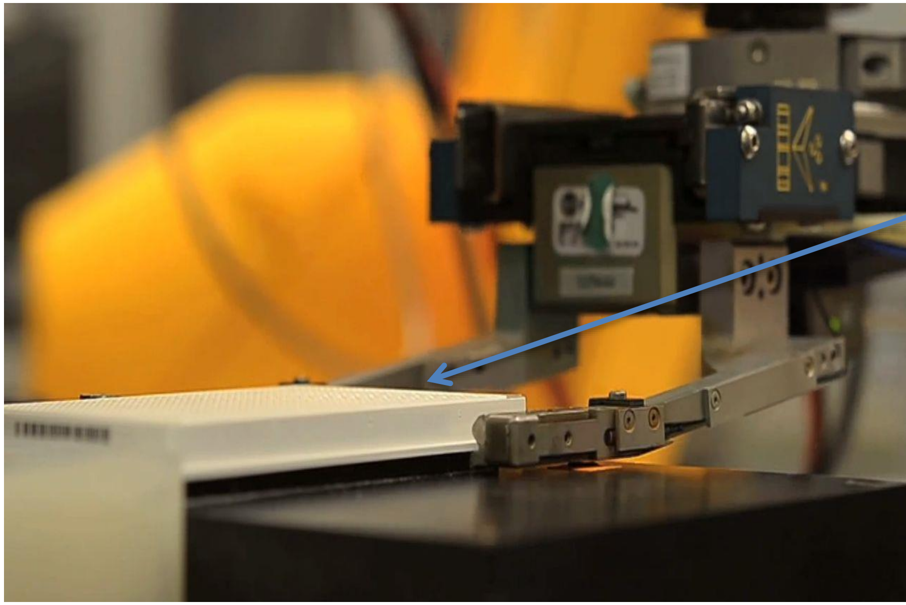


Close up of a few cells

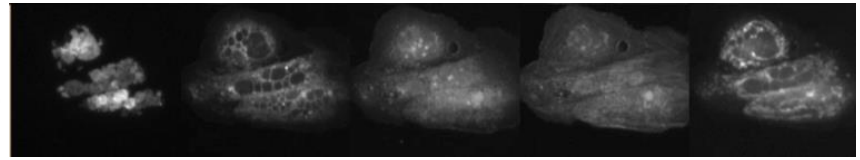
Biologist

She can tell difference

Phenotypic High Content Screening: Robotics at Scale



One well (close up)



Small screen (by industry standards)

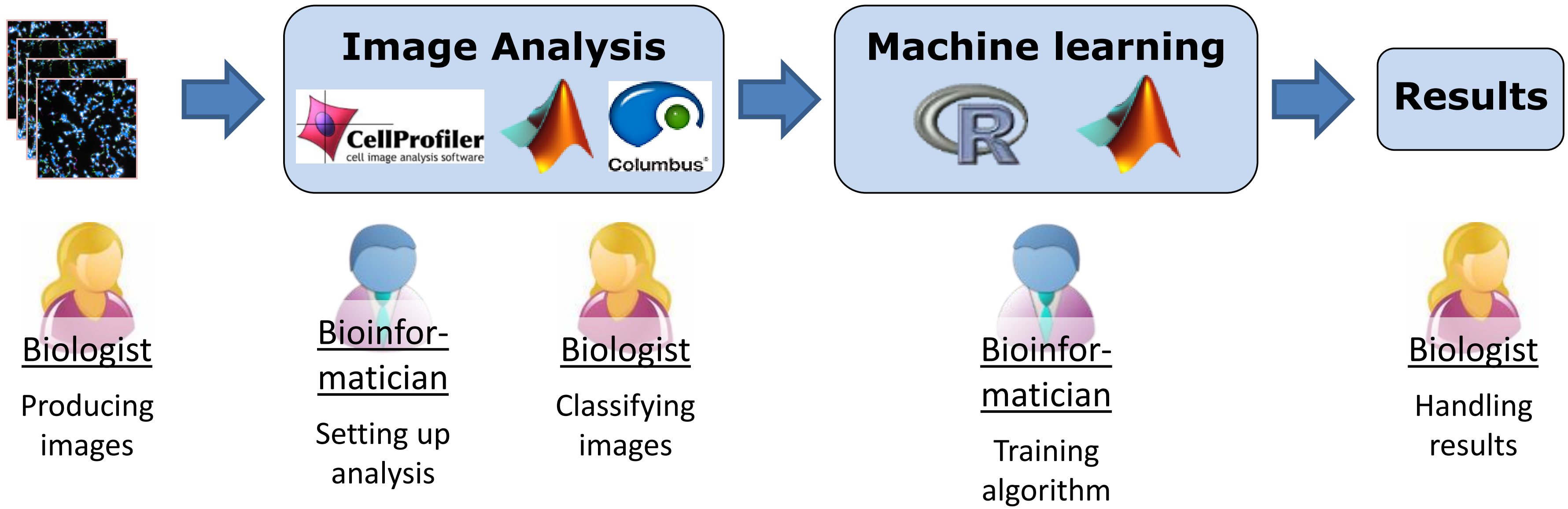
- 20 plates
- 384 wells
- 900 cells per well

~7 million cells in the screen (350 GB)

➔ need for an automatic approach…

Classical Image Analysis workflow for HCS

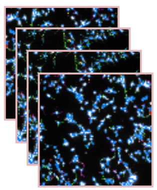
Classical HCS workflow



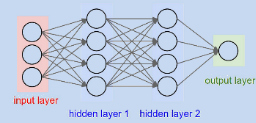
Slide Credit: Genedata AG Basel

Deep Learning-based workflow of HCS

Deep Learning-based HCS workflow



Train & Apply CNN



Results

Biologist

Producing images

Biologist

Classifying phenotypes

Biologist

Handling results

- No time-consuming tuning of image analysis algorithms
- No scripting expertise required
- Single convenient process from start to finish

- Classify training data by simple drag-and-drop
- No tuning of experiment protocols to fit image analysis needs
- Reduction of project times !!

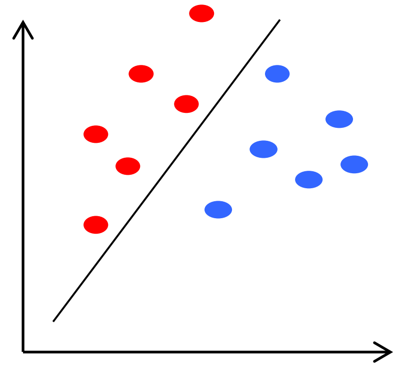
Slide Credit: Genedata AG Basel

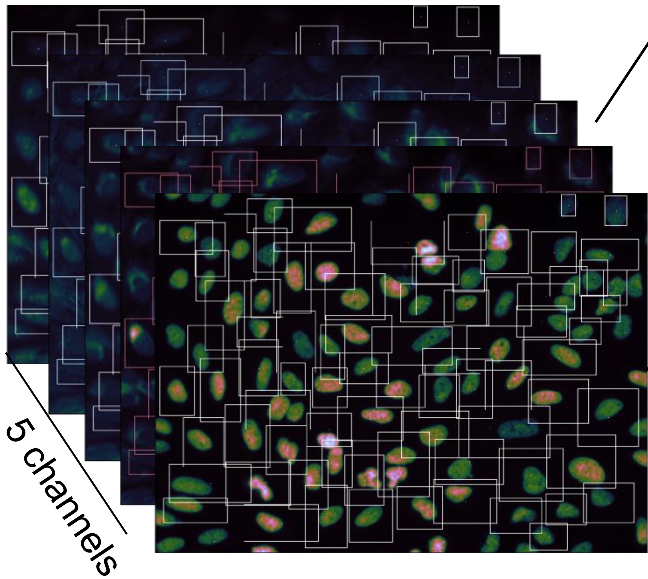
Comparison

SVM, LDA, RF (baseline)

Extracted features (a 396 dimensional vector for a cell)

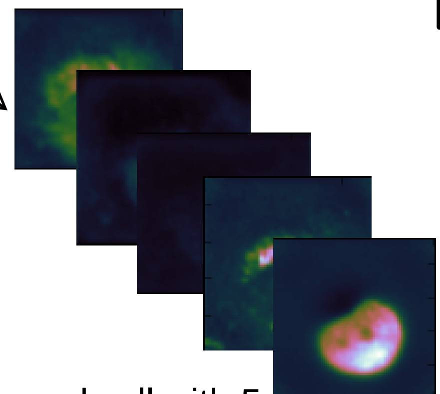
(1.2, 0.02,…,42)





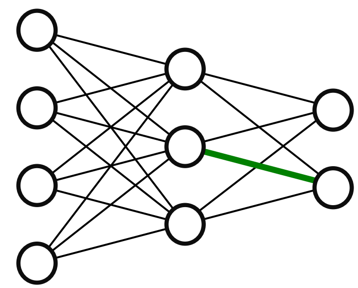
The 5 channels correspond to different stainings.

CNN (new)



Cropped cell with 5 channels as input for CNN.

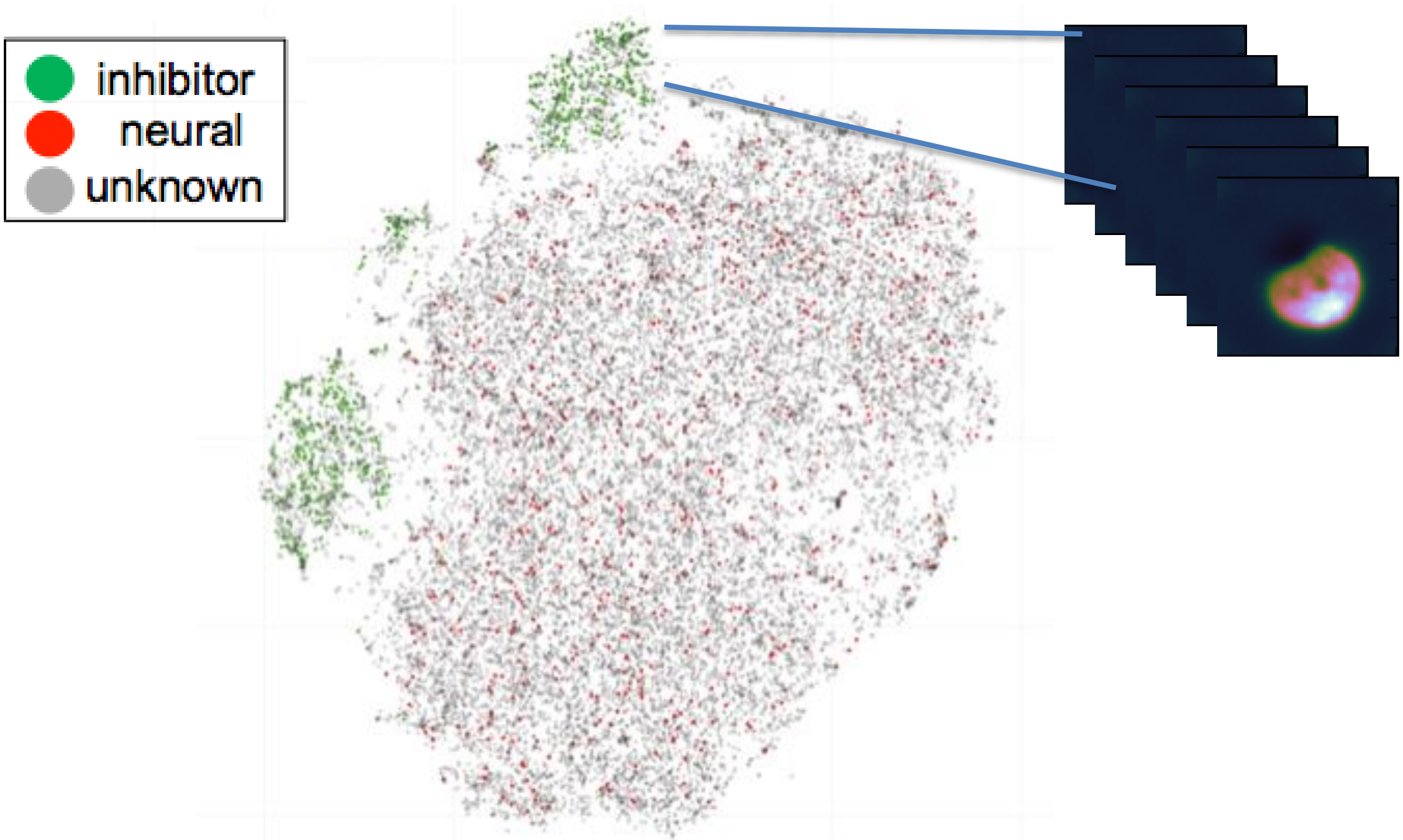
Feature extraction using a CNN



- Current Project with Genedata AG funded by the Swiss Government (CTI, 300+ kCHF)
- Dürr, O., and Sick, B. “Single-cell phenotype classification using deep convolutional neural networks”. *Journal of biomolecular screening* 21, 9 (2016), 998-1003

Unsupervised Learning for fast creation of training set

- We can help even more by pre sorting
  - Use a network trained on ImageNet (cats&dogs) VGG16
  - Feed images of cells into network
  - Use intermediate vector (FC7) as input to tSNE

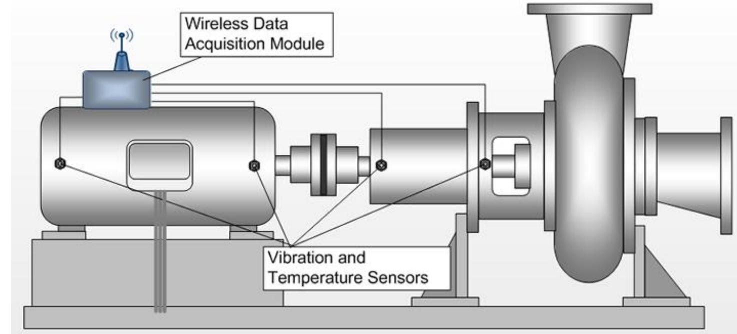


Biologist

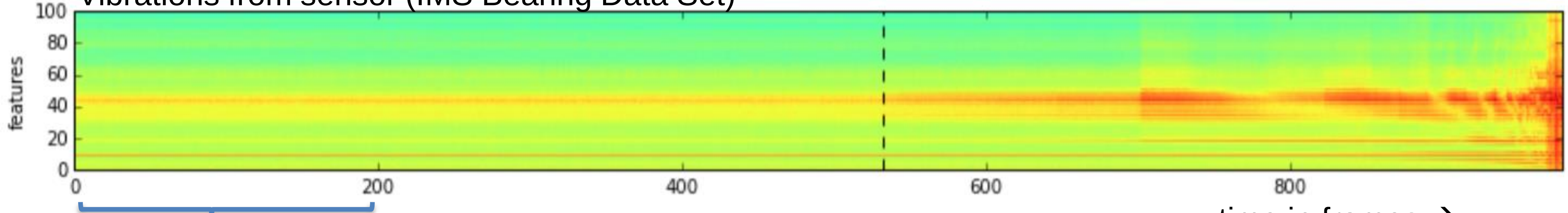
She can tell difference on a batch

Case Studies III-IV

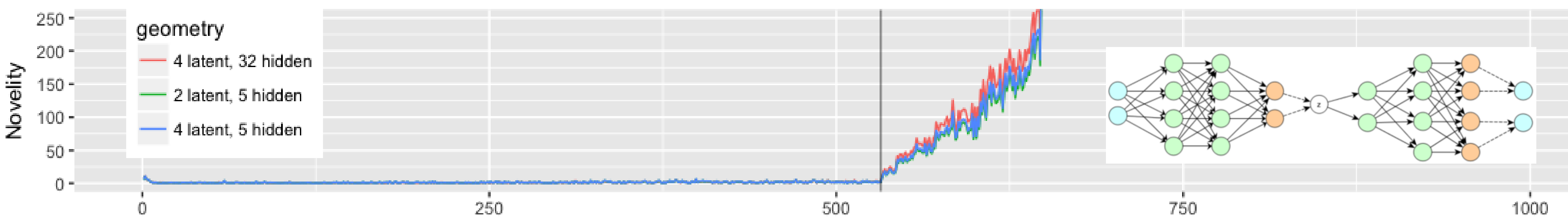
Condition Monitoring



Vibrations from sensor (IMS Bearing Data Set)



Deviation from normal condition infered by a variational auto encoder

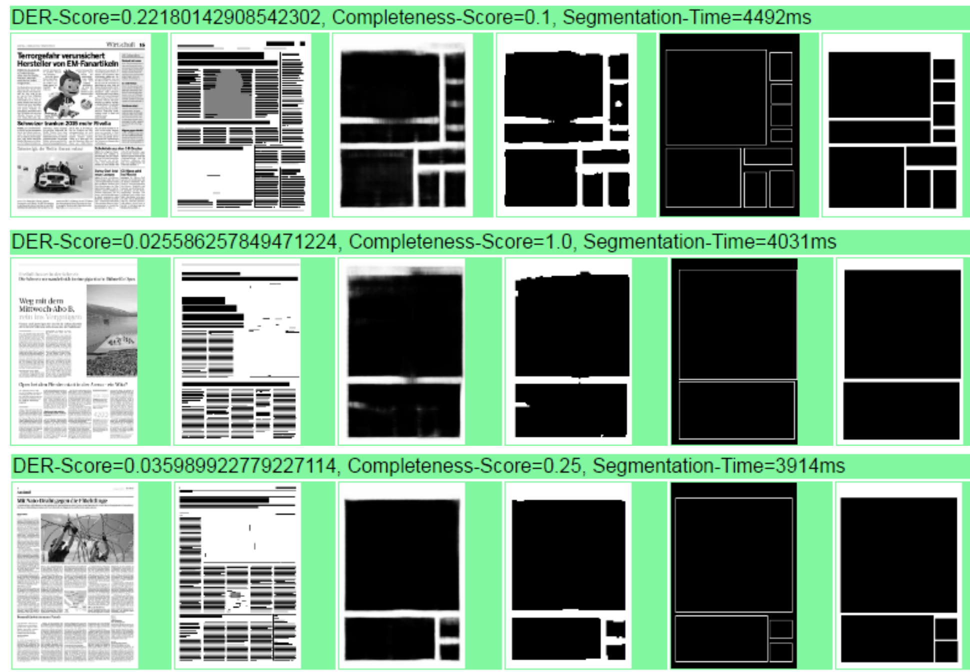


- CTI Project Data Driven Condition Monitoring DaCoMo (CTI, 350+ kCHF)
- Stadelmann, T; Tolkachev V; Sick, B; Stampfli, Dürr, O; J. „Beyond ImageNet - Deep Learning in Industrial Practice“ submitted to *Applied Data Science* Braschler, M; Stadelmann, T.; Stockinger, K. (Eds.) Springer

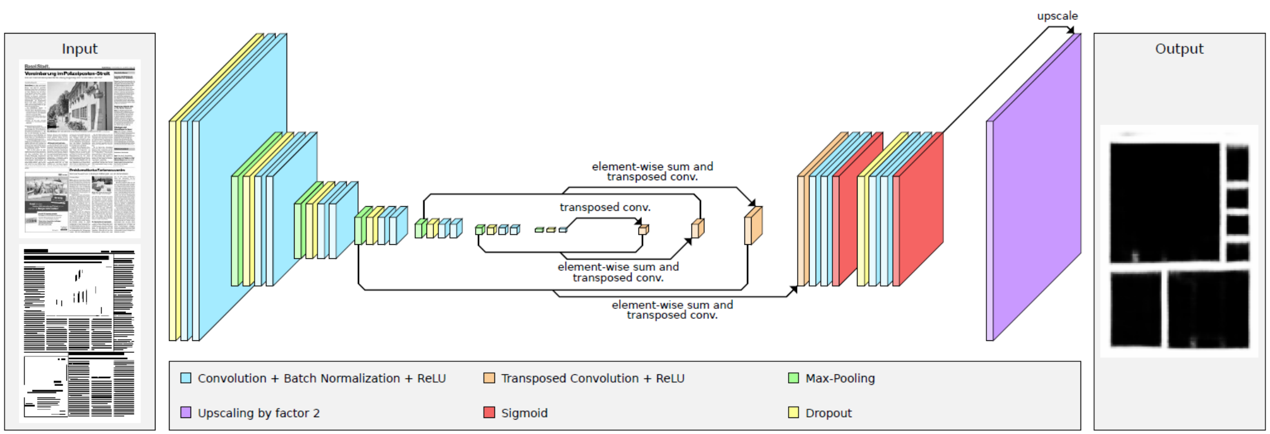
Learning to segment:
Vision-based newspaper article segmentation

Task:

Results:



Solution:



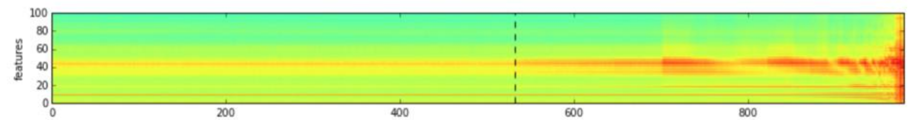
Zürcher Hochschule
für Angewandte Wissenschaften

3

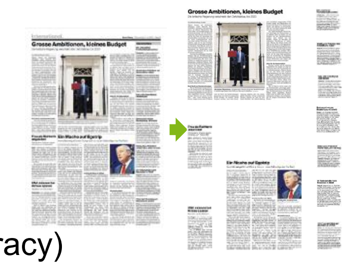
DL using limited resources, and other advice

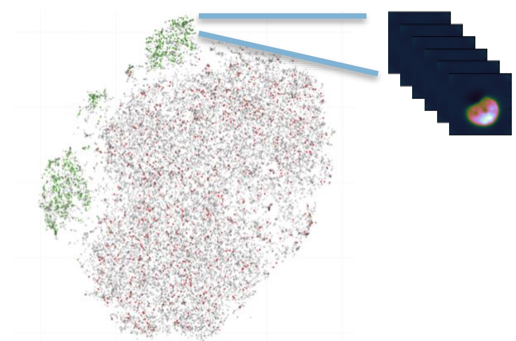
Beyond ImageNet?
We've successfully trained non-classical DL models to …

- **Learn to detect novelties** / anomalies
  – Task: predictive maintenance
  – Approach: autoencoders
  – Training data: ca. 400 time points per machine
  – Satisfaction: medium (worked well for faults humans can recognize; **fails for faults humans don't see coming**)



- **Learn to segment**
  – Task: segmentation of newspaper pages into articles
  – Approach: fully convolutional neural networks
  – Training data: ca. 500 fully labeled pages, ca. 5'000 partially labeled pages
  – Satisfaction: high (much **better than all other automatic approaches**; ca. 70% "felt" accuracy)





- **Learn to use existing networks**
  – Task: create labeled data for training
  – Approach: use existing network as feature extractor, then cluster
  – Satisfaction: high (faster human interaction for **creating training data**)

Tips for working with limited data

- **Transfer learning**
  - Use pre-trained networks designed for a “close enough” task
    (e.g. VGG-16 for image classification)

- **Trainable architectures**
  - Use architectures like Inception or ResNet that adapt their complexity to the available data
  - Use network compression (less parameters → less data needed)

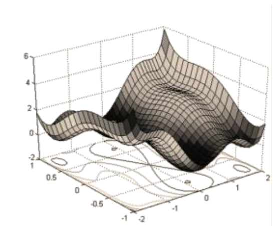
- **Data augmentation**
  - Provide variants of original data that
    (a) you can create randomly **on the fly**
    (b) resemble distortions / alterations **relevant and realistic** in practice

- **Unlabeled data**
  - Employ semi-supervised learning
  - Use high-level features created by a first net to do a clustering / t-SNE embedding
    → This allows to label lots of data after a short inspection

Important lessons learned

- **A good baseline**
  – For a new use case, start from an easy & well understood baseline model (i.e., one closely resembling a published architecture and task)
  – **Increase** the **complexity** of the architecture **slowly**.

- **A suitable loss function**
  – Ensure to provide a loss function which **really describes the problem** to be solved
  – Especially useful if the task is not classification (e.g., clustering)



- **Debugging**
  – **A first DL model** on a completely new task and data set **usually does not work**
  – Options:
    - **Hand-calculating** training **equations** for toy examples → bugs e.g. in the loss function?
    - **Visualizing** the pre-processed **data** → bug in data loading?
    - **Visualizing** expected **loss** values → does it learn at all?
    - **Inspecting misclassified** training examples → get intuition into what goes wrong

Conclusions

- **Solid technical foundations** and a **growing base of use cases**
  ➔ the technology is ready to apply to new areas (in pattern recognition)
- DL tech **transfer is really fast**: ~3 month from publication to industry
- **Success** (on non-traditional use cases) depends on experience & experiments

How to find us

- Dr. Thilo Stadelmann
- Head of ZHAW Datalab, Vice President SGAICO, Board Data+Service
- thilo.stadelmann@zhaw.ch, @thilo_on_data
- 058 934 72 08
- www.zhaw.ch/~stdm

- Dr. Oliver Dürr
- Deputy head of ZHAW Datalab, Senior Lecturer Statistical Data Analysis
- oliver.duerr@zhw.ch
- 058 934 67 47
- www.zhaw.ch/~dueo

datalab www.zhaw.ch/datalab

swiss group for artificial intelligence and cognitive science

Swiss Alliance for Data-Intensive Services

APPENDIX

CNN for HCS

Table 1. Architecture of the Convolutional Neural Network.

Layer Description	No. of Images/Feature Maps × Their Dimensions	No. of Weights
Input	5 × 72 × 72	
Convolution (3 × 3)	32 × 70 × 70	32*9*5
Convolution (3 × 3)	32 × 68 × 68	32*9*32
Max pooling (2 × 2)	32 × 34 × 34	
Convolution (3 × 3)	64 × 32 × 32	64*9*32
Convolution (3 × 3)	64 × 30 × 30	64*9*64
Max pooling (2 × 2)	64 × 15 × 15	
Convolution (3 × 3)	128 × 13 × 13	128*9*64
Convolution (33)	128 × 11 × 11	128*9*128
Max pooling (2 × 2)	128 × 6 × 6	
Fully connected	200	200*128*6*6
Fully connected	200	200*200
Fully connected	50	50*200
Output	4	4*50

Newer Versions use dropout but no batch-norm

More non-traditional DL applications
…beyond image classification

- **Learn to cluster**
  - Task: cluster 1s long speech utterances into an unknown number of unknown speakers
  - Approach: CNN/RNN to extract embeddings for subsequent hierarchical clustering
  - Training data: ca. 20s from 100 speakers
  - Satisfaction: high (achieved state of the art, doesn't work end-to-end yet)